DOCUMENT RESUME

ED 065 590 TM 001 859

AUTHOR Kennedy, Beth T.
TITLE The Role of Criterion-Referenced Measures within the Total Evaluative Process.
PUB DATE Apr 72
NOTE 4p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, Illinois, April 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Criterion Referenced Tests; *Evaluation Methods; *Instructional Programs; *Test Reliability; *Test Validity
IDENTIFIERS *Southwest Educational Development Lab

ABSTRACT

Issues related to the evaluation of instructional programs developed under the auspices of the Southwest Educational Development Laboratory are briefly discussed. The Laboratory develops criterion-referenced tests which form an integral part of each instructional program. The importance of examining the reliability and validity of these tests is noted. (DB)

ED 065590

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

THE ROLE OF CRITERION-REFERENCED MEASURES WITHIN THE TOTAL EVALUATIVE PROCESS

Beth T. Kennedy, SEDL

The Southwest Educational Development Laboratory is committed to identifying educational problems, particularly those related to the culture and economy of the Southwest, and to assure the attainment of educational improvements through innovations directed toward change in educational practice. The problem focus is intercultural education, emphasizing the cultural richness of the diverse population groups of our region through the arts, through language study and through better understanding of ethnic heritages.

Much of the work within the Laboratory is devoted to the design, evaluation and development of curricula and teaching strategies to provide those experiences which maximally develop children's potential. The major concern being expressed in this paper and those which follow deals directly with issues related to the evaluation of instructional programs developed under the auspices of the Laboratory.

All instructional programs are developed from specified objectives. These objectives range from general program goals to more specific goals for individual lessons. The objectives within these programs form a hierarchy; lessons are designed to systematically integrate abilities and skills and progressively increase the complexity of mental abilities required to successfully complete a curriculum.

While the Laboratory is interested in the extent to which these instructional programs produce general changes in behavior, as measured by norm-referenced measures of achievement, our primary concern is in ascertaining the extent to which the instructional programs produce more specific changes in behavior.

TM 001 859

Criterion-referenced measures provide data regarding the child's ability to demonstrate mastery of established criterion behaviors. Different levels of criterion referenced measures are developed to more completely assess various levels of behavior. Objectives which pertain to individual lessons are assessed through unit tests. Unit test items are written for each lesson and assess recall of specific skills or knowledge directly pertaining to that lesson. More general instructional goals are stated in the form of objectives which reflect higher mental abilities (e.g., application, analysis, synthesis, and evaluation). These broader program objectives are assessed through mastery test items. Again, individual items are written to assess mastery of these broader objectives for each program. Thus, criterion-referenced measures are currently being employed to provide learner performance data. As the name implies, criterion-referenced measures are designed to measure the child's ability to master specific criterion behaviors. The behaviors falling within the focus of evaluation are defined by specific criteria stated in the instructional objectives of a program. The results of these performance measures are used to evaluate the performance of a group with respect to an absolute standard, thereby negating the need for interpersonal comparisons. In addition to charting specific growth curves, data from criterion-referenced measures are important in evlauating and revising the program content, in evaluating the efficacy of instructional sequences, and in evlauating competing instructional programs and products.

Unfortunately, the profession has not agreed on a uniform definition for criterion-referenced measures. Glaser (1963) defines criterion-referenced measures as those which "depend upon an absolute standard of quality" as opposed to norm-referenced measures which "depend upon a relative standard." Livingston (1970) refers to a criterion-referenced measure as "any test for which a criterion score is specified, without reference to the distribution

of scores." Kriewall (1969) defines a criterion-referenced test as one which is constructed to provide proficiency measures relative to a specific class of problems according to an item sampling model which requires that the test items be homogeneous in difficulty. Ivens (1970) defines a criterion-referenced measure as one which is "comprised of items keyed to a set of behavioral objectives." Glaser and Nitko (1970) clarified Glaser's earlier definition by stating, " a criterion-referenced test is one that is deliberately constructed to yield measurement that is directly interpretable in terms of specified performance standards." The Laboratory develops criterion-referenced tests from specified objectives which are directly interpretable in terms of the specific performance standards stated in those objectives. Therefore, our conception of criterion-referenced measures is similar to those definitions provided by Ivens and Glaser and Nitko. In developing criterion-referenced measures, the Laboratory recognized the corresponding need to estimate the reliability and validity of these measrues. Laboratory specialists in curriculum as well as those in measurement often expressed reservations about the quality of criterion-referenced items. Therefore, it was important to more fully examine the reliability and validity of these items to determine our conficence in basing decisions on data supplied from these measures.

However, there was another major reason for our examining the reliability and validity of these measures. All instructional programs eventually will be available commercially. In keeping with the standards of the Laboratory, it is important to fully examine the characteristics of the products being produced, in an effort to provide viable instructional programs. Therefore, in that criterion measures form an integral part of each instructional program, it is important to examine the reliability and validity of these measures.

We will be discussing issues relevant to our attempts to examine the reliability and validity of the criterion-referenced measures produced by the Laboratory.

BIBLIOGRAPHY

Glaser, R. Instructional technology and the measurement of learning outcomes. American Psychology, 1963, 18, 519-521.

Glaser, R., & Nitko, A. Measurement in learning and instruction. in R.L. Thorndike (Ed) Educational Measurement, Washington, D.C.: American Council on Education, 1971.

Ivens, S. An investigation of item analysis, reliability, and validity in relation to criterion-referenced tests. Doctoral dissertation, Florida State University, 1970.

Kriewall, T. Applications of information theory and acceptance sampling principals to the measurement of mathematics instruction. Doctoral dissertation, University of Wisconsin, 1969.

Livingston, Samuel A., The Reliability of Criterion-Referenced Measures. Baltimore, Maryland, The Center for the Study of Social Organization of Schools, The Johns Hopkins University, Report No. 73, July, 1970.